Against the Grain

Volume 21 | Issue 6 Article 45

December 2009

ATG Interviews Lotfi Belkhir

Martha Whittaker
George Washington University, mwhittaker@gelman.gwu.edu

Recommended Citation
Whittaker, Martha (2009) "ATG Interviews Lotfi Belkhir," *Against the Grain*: Vol. 21: Iss. 6, Article 45.
DOI: https://doi.org/10.7771/2380-176X.2390

Op Ed
from page 32

As to the second point, I'm arguing only that intrinsic quality should not be the only (or, in some cases, even the primary) criterion for acquisition. The primary criterion should be *usefulness*, and usefulness is determined by the library's mission and the needs of its patrons. Bad books can be very useful indeed, and *Arming America* strikes me as an eminent example of such a book.

By providing *Arming America* to its patrons, does the library run a risk that they will come away from it with dangerously incorrect information? Absolutely. But this raises a deeper issue with which our profession has grappled for more than a hundred years. Giving people access to information is, by its nature, a risky business. If we believe that knowledge is powerful, then we have to accept that it is therefore also dangerous. When we expose people to ideas we are not only fostering understanding and empathy and the broadening of minds, but are also running the risk that wrongheaded and mean and chauvinistic ideas will sprout and take hold. However, it seems to me that as librarians we have no choice but to take that risk. The alternative is actually rather awful to contemplate.

A Response from Steve McKinzie

by **Steve McKinzie** (Library Director, Corriher-Linn-Black Library, Catawba College, Salisbury, NC 28144; Phone: 704-637-4449) <smckinzi@catawba.edu>

Mr. Anderson makes his case eloquently and persuasively. In a spirit of friendly exchange, I counter briefly. In the final analysis, my recommendation to get rid of *Arming America* hinges on a simple distinction that bears repeating — a distinction about scholarly books that contain inaccuracies. Simply put, I draw a line. I suggest there is a huge difference between skewed historical analysis and deliberately falsified research — between a historian who may have a jaundiced perspective and one who knowingly chooses to deceive — between an honest scholar who misreads his data and dishonest one who deals in blatant misrepresentation of his sources. I charge that in every case, *Arming America* falls on the latter side of this line of distinction and consequently merits no place in a scholarly collection.

Let's be candid. If any of us knew then what we know now about *Arming America*, would we have purchased the title? I think not. And there something else we should consider. If we eagerly discard older historical monographs and outdated research in the interests of saving precious shelf space, should we not also willingly jettison a title based on what we have learned about the utter dishonesty of its approach — indeed the utter dishonesty of the research on which it is based?

Don't get me wrong. I appreciate **Mr. Anderson's** analysis, and there is likely much on what we agree. But as for *Arming America*, I remain respectfully unconvinced. I still say throw it out.

ATG Interviews Lotfi Belkhir

Founder and CEO of Kirtas Technologies

by **Martha Whittaker** (Director, Content Management, Gelman Library System, George Washington University) <mwhittaker@gelman.gwu.edu>

ATG: Let's start by learning about your background leading up to the founding of Kirtas Technologies.

LB: I completed my PhD in physics in 1993, and in 1995 I left the academic community to join **Xerox** as a research scientist. My five-year career at **Xerox** spanned R&D, product development, corporate strategy and corporate ventures. I started the **Automatic Book Scanner** project while an executive at the **Xerox Venture Lab** in Palo Alto, CA in late 2000. In May 2001, I left **Xerox** with an exclusive license to the technology I developed, and started **Kirtas** in June 2001.

ATG: Your Website has the engaging tag line "Moving knowledge from books to bytes." Tell us about Kirtas Technologies, Inc. — how it was born and the path of its development.

LB: Kirtas owes its existence and its purpose to four global trends that are and will continue to affect our lives for many years to come: Digitization, Globalization, Knowledge-driven economy and the Internet. These trends are also intimately intertwined. On the other hand, back in 2000, while all forms of content, communication and entertainment were going digital, the accumulated knowledge of humankind from the last 1,000 years was still largely held captive in the analog world by the covers that bind it. I'm of course talking about the billions of books that lay on the library shelves of thousands of libraries around the world. The reason being that there was no technology available to enable the massive, rapid, high-quality and cost-effective conversion of all that content "from books to bytes." So I set out to change that and develop that dearly needed solution. Started in June 2001, **Kirtas** went on to develop the first generation model, the **BookScan APT 1200**, which we launched in August 2004. Some of our earliest customers were **Logo Bible Software**, **EBSCO**, **Northwestern University** and **Rochester Public Library**. We also opened a service bureau to provide digitization services a few months earlier, and our earliest customers were **Atypon**, **University of Michigan**, as well as **EBSCO Publishing**.

ATG: Who are some of your clients? Libraries? Publishers? Others?

LB: Today, **Kirtas** products are present in more than 30 countries with over 400 customers around the world. Our client list includes some of the most prestigious names in the academic, research, government, corporate, publishing and non-profit library world such as **Yale U**, **Cornell U**, **Emory U**, **John Hopkins U**, **Emory U**, **Novartis**, the **Air Force**, the **United Nations Organization**, the **British Library**, **Cambridge University Press**, **Hong Kong U**, **Government of Canada**, **McGill U**, **Polytechnic Institute of St Petersburg**, **Yeltsin Presidential Library**, and the list goes on.

ATG: Why are the terms "scanning" and "digitization" not to be used interchangeably?

LB: Scanning is usually understood as the process of capturing a digital image of a document, while digitization means the process of converting the content of that document into a readily usable digital file. While obviously digitization requires scanning as a first step, it also requires additional post-processing steps which, in the case of complex documents such as books, are usually far more challenging and technically sophisticated than the scanning step.

ATG: What do you mean when you talk about the "three pillars of digitization?"

LB: Kirtas introduced the concept of the "three pillars of digitization" to describe in more concrete terms how different digitization is from scanning, and what are the three fundamental components of digitization that in practice must be delivered by the digitization process in order to ensure the longevity, the interoperability and the repurposing of the digitized assets.

ATG: How do you define "quality" as it is applied to digitization projects?

LB: Defining "quality digitization" has been a thorny issue for librarians since the "Making of America" project by **Cornel University** and **U of Michigan**. Too often it centers on DPI, output format, OCR accuracy, full color vs. bitonal, etc. Needless to say that with improving technology, these quality specifications have became a moving target. But

continued on page 36

more importantly, none of these specifications really offer a true definition of quality that can become a standard "umbrella" definition from which a well thought-out and logical set of specifications will flow. Hence, quality today is still "in the eye of the beholder."

At **Kirtas**, we define quality digitization as follows:

> Quality Digitization is the creation of digital assets that: 1) can survive the passage of time; and 2) can be programmatically repurposed to meet the new and existing access needs of today and in the future.

We believe that from this high-level definition we are able to infer a complete set of detailed specifications that will meet the requirements of both digital preservation and multi-purpose access while leveraging our initial investment.

ATG: Please talk about what you refer to as the "cost and opportunity cost of digitization."

LB: Digitization is often viewed as an expensive endeavor. So to respond to the increasing demand for electronic access by their patrons within the limitations of their ever-shrinking budget, libraries feel often forced to "cut corners" when it comes to quality. The common wisdom is that quality must be sacrificed for quantity when it comes to large-scale digitization because it's assumed that quality digitization will involve a much higher cost. Our experience is that large-scale quality digitization when done within an efficient workflow, while not free, doesn't cost that much more than its poor counterpart. In fact, on a "total cost" basis, the evidence points to the contrary. Furthermore, there's an enormous long-term "opportunity" cost for not doing quality digitization, especially when it's on a large scale. That opportunity cost includes the (i) perception that it's a "done task" and hence eliminates any future financing opportunity to do it right, (ii) the detrimental impact on the institution's reputation, (iii) the inability to repurpose that content down the road to adapt to the rapidly emerging formats (e.g, ePub for eBook reading devices), and needs of their patrons, (iv) the high error rate inherent to those low-quality digitization programs that makes them unsuitable for scholarly research. In short, I believe that the short-term gains (if any) of having poorly digitized collections are dwarfed by those long-term opportunity costs.

ATG: Why should librarians planning digitization projects be concerned with workflow, and what are the steps that should be a part of the planning process?

LB: Because digitization is so much more than scanning, the cost of the scanning tends to be a small part, usually between 10-15% of the total cost of digitization. This total cost must include not only the obvious costs of post-processing, but also that of prioritization, selection, pulling, inspection for rejects, file management, quality control, data entry, data storage, re-shelving, and so on. A well designed and efficient workflow process is hence critical to achieve a large-scale, high-quality and cost-effective digitization program.

ATG: What should we be doing to insure interoperability and sustainability of our digitization projects?

LB: Interoperability requires common standards for image quality, image formats, and metadata. Sustainability on the other hand requires the kind of quality and file formats that lend themselves to a programmatic conversion to other formats and uses that we may or may not foresee today. To illustrate this key difference, let's assume that a consortium of libraries agrees on a joint and distributed digitization program with an agreed upon and well-defined set of specifications about the image quality, image format, OCR accuracy level, and metadata output. As long as they all abide by those standards, then their separately digitized assets should be interoperable. On the other hand, say they want to convert down the road their files to print-ready PDF's for print-on-demand. Depending on how they agreed to process their images during their digitization program will determine whether they can programmatically (and hence at low cost) convert those digital assets, or whether they would have to manually convert every book (and hence at a prohibitive cost). The same goes if they want to create an ePub file. The accuracy level of their OCR, and the level of completeness of their structural metadata will also determine whether a clean and usable ePub is possible programmatically or not. If the answer is no in both cases, then those digital assets, while interoperable are not sustainable and hence will become quickly obsolete.

ATG: You have a print-on-demand service KirtasBooks.com. What is the business model for this service and what distinguishes it from other POD?

LB: Kirtasbooks.com business model differentiates itself by the concept of "Comprehensive Access." Every title we have on our one million-record database is available either on-demand, or already digitized. Comprehensive Access means that once digitized, that title is available for free online reading, as a case-bound or paperback POD, as a fully searchable download, or as part of a research collection using the best of breed technology from ebrary. Today, no other company in the world, whether it's **Amazon**, **Google**, **Ingram** or other offers that kind of access model.

ATG: I know you recently announced a partnership with OCLC. Could you tell us more about that?

LB: Our exciting partnership with **OCLC** allow us to update **Wordcat** records each time we digitize a title from our database of Digitize-on-Demand (DoD) records. That digitized title will then become visible, via a persistent hyperlink, to any **OCLC** member around the world.

ATG: What are some other big projects in the works that you can tell us about?

LB: All our projects center around the digitization and distribution of library content, with a key focus on quality in the sense of longevity, interoperability and sustainability. Based on our customers' feedback, we continue to develop hardware and software products that on one end of the spectrum enable the conversion of special collections material, and on the other end of the spectrum enable the clean and painless repurposing of digitized material into emerging formats such as the ePub format.

ATG: In your opinion, who are the major digitization players now and what will the industry look like in five years?

LB: With **Microsoft's** exit from this arena last year, the three other major players, besides **Kirtas**, are **Google**, **Amazon** and the **Internet Archive**. **Google** is by far the largest player today, although their output falls far short from qualifying as "digitization." **Amazon** does a good job in its digitization of relatively recent and copyrighted material, but all its scanning is destructive. The **Internet Archive** has made commendable efforts in scanning manually over a half-million titles and making them available on its archive, but there again because of their narrow focus on scanning only and rudimentary quality control process, there will need to be extensive additional investment to make their scanned titles interoperable and repurposable. I believe that five years from now, the industry will have finally settled on a thorough set of standards of digitization that will enable interoperable content to be accessed in multiple ways. I see new international partnerships forming between libraries, governments, and corporations forming around the preservation, access and sharing of information and knowledge in ways only possible through the digital media. I see new technologies emerging spurred by new opportunities that will be created and stimulated by all the digitized content. I see the Web 3.0 taking full shape and an unprecedented level of knowledge and information sharing taking place.

ATG: What message do you have for new people just entering the library and information services profession?

LB: With the information explosion and the increasing need for researching and accessing vast amounts of information, there's also a commensurate need for a new breed of library services professionals that will adapt to the rapidly changing landscape. They must learn to leverage the latest technologies to enhance not only the quality of their services but also the efficiency with which they deliver those services. Indeed digitization, when done right, is one such enabler of quality and efficiency. However to achieve its full potential, digitization needs to be embraced and owned directly by the library community instead of other organizations who share neither its know-how, nor its agenda.

ATG: What one thing would you like readers of Against the Grain to think about when they think of Kirtas?

LB: A partner who's passionately dedicated to "moving knowledge from books to bytes."